Supporting Information

Hong et al. 10.1073/pnas.1515982112

Features of Individual Animals

For each frame t of the recorded video, an ellipse was fit to each animal (e.g., resident or intruder), characterized by the five parameters $\{x^n(t), y^n(t), l^n(t), s^n(t), \theta^n(t)\}$, $n \in \{R$ (for the resident), I (for the intruder)$\}$, where (x^n, y^n) are the Cartesian coordinates of the ellipse centroid relative to the bottom left corner of the home cage; l^n is the length of the major axis; s^n is the length of the minor axis; and θ^n is the body orientation in degrees. Sixteen features were extracted from the 10 ellipse parameters (5 for each animal):

Feature 1. Speed of forward motion of the resident centroid:

$$V(t) = \left\| x^R(t+4) - x^R(t-4),\, y^R(t+4) - y^R(t-4) \right\| \cdot \cos\left(\theta^R - \varphi^R\right),$$

where $\|\cdot\|$ is the Euclidean norm and φ^R is the direction of motion of the centroids.

Feature 2. Speed of forward motion of the intruder centroid:

$$V(t) = \left\| x^I(t+4) - x^I(t-4),\, y^I(t+4) - y^I(t-4) \right\| \cdot \cos\left(\theta^I - \varphi^I\right),$$

where $\|\cdot\|$ is the Euclidean norm and φ^I is the direction of motion of the centroids.

Feature 3. Change of the body orientation of the resident (Fig. S7*B*):

$$\Delta\theta^R = \mathrm{mod}\left(\theta^R(t) - \theta^R(t-1), 360\right).$$

Feature 4. Change of the body orientation of the intruder (Fig. S7*C*):

$$\Delta\theta^I = \mathrm{mod}\left(\theta^I(t) - \theta^I(t-1), 360\right).$$

Feature 5. Area of resident ellipse:

$$A^R(t) = \pi \frac{l^R(t)}{2} \frac{s^R(t)}{2}.$$

Feature 6. Area of intruder ellipse:

$$A^I(t) = \pi \frac{l^I(t)}{2} \frac{s^I(t)}{2}.$$

Feature 7. Aspect ratio between lengths of the major and minor axis of resident ellipse:

$$R^R(t) = \frac{l^R(t)}{s^R(t)}.$$

Feature 8. Aspect ratio between lengths of the major and minor axis of intruder ellipse:

$$R^I(t) = \frac{l^I(t)}{s^I(t)}.$$

Feature 9. Height of the highest point along the major axis of resident ellipse: Given depth sensor reading $z^R(p_x, p_y, t)$ at pixel (p_x, p_y) in frame t, define nine evenly spaced points along the major axis

$$\left(s_x(i), s_y(i)\right) = \left(x^R(t), y^R(t)\right) + \frac{i}{5}\left(\frac{l^R(t)}{2}\cos\left(\theta^R(t)\right), \frac{l^R(t)}{2}\sin\left(\theta^R(t)\right)\right)$$

for $i \in [-4,4]$.

Compute the average depth $Z_i^R(t)$ within a square region of width $r = l^R(t)/10$ centered at each point:

$$Z_i^R(t) = \frac{1}{2r}\sum_{p_y = s_y(i)-r}^{s_y(i)+r}\left(\frac{1}{2r}\sum_{p_x = s_x(i)-r}^{s_x(i)+r} z^R\left(p_x, p_y, t\right)\right),$$

then take the maximum, $H^R(t) = \max(\{Z_i^R(t)\})$.

Feature 10. Height of the highest point along the major axis of intruder ellipse:

$$H^I(t) = \max\left(\left\{Z_i^I(t)\right\}\right),$$

where $Z_i^I(t)$ is defined as in feature 9.

Feature 11. Relative angle between body orientation of the resident $\theta^R(t)$ and the line connecting the centroids of both animals (Fig. S7*D*):

$$\Phi^R(t) = \left\lfloor \mathrm{mod}\left(\theta^R(t) - \mathrm{atan2}\left(\frac{r_y(t)}{r_x(t)}\right), 360\right)\right\rfloor,$$

where $\lfloor x \rfloor = \min(x, 360 - x)$ and $(r_x(t), r_y(t)) = (x^R(t) - x^I(t), y^R(t) - y^I(t))$.

Feature 12. Relative angle between body orientation of the intruder $\theta^I(t)$ and the line connecting the centroids of both animals (Fig. S7*E*):

$$\Phi^I(t) = \left\lfloor \mathrm{mod}\left(\theta^I(t) - \mathrm{atan2}\left(\frac{r_y(t)}{r_x(t)}\right) - 180, 360\right)\right\rfloor,$$

where $\lfloor x \rfloor$ and $(r_x(t), r_y(t))$ are defined as in feature 11.

Feature 13. Distance between the two animals (Fig. S7*F*):

$$D(t) = \left\| x^R(t) - x^I(t),\, y^R(t) - y^I(t) \right\| - \left\| c^R(t) \right\| - \left\| c^I(t) \right\|,$$

where $c^n(t) = [l^n(t) \cdot \sin(\Phi^n(t)), s^n(t) \cdot \cos(\Phi^n(t))]$, $n \in \{R, I\}$ is the length of the semiaxis of the ellipse along the line connecting

the centroids of both animals, and $\Phi^n(t)$ is defined as in features 11 and 12.

Feature 14. Distance between the two animals divided by length of the semiaxis of the resident along the line connecting the two centroids:

$$R_D(t) = \frac{D(t)}{c^R(t)},$$

where $D(t)$ and $c^R(t)$ are defined as in feature 13.

Feature 15. Ratio between sizes of resident ellipses and intruder ellipses:

$$R_S(t) = \frac{A^R(t)}{A^I(t)},$$

where $A^n(t)$ is defined as in features 5 and 6.

Feature 16. Pixel changes from side-view video frames: Given monochrome light intensity $f(p_x, p_y, t)$ at pixel (p_x, p_y) in frame t,

$$\Delta P(t) = \sum_{p_x,p_y} \left(f(p_x,p_y,t) - f(p_x,p_y,t-1) \right)^2 \Big/ \sum_{p_x,p_y} \left(f(p_x,p_y,t) \right)^2.$$

Features 17–26. Smoothed features computed by averaging other extracted features over a 0.367-s window (± 5 frames at 30 Hz, for 11 frames total). Smoothing was applied to features 1, 2, 5, 6, 7, 8, 9, 10, 13, and 16 to create features 17–26, respectively.

Feature 27. Smoothed feature computed by averaging feature 16 over a 5.03-s window (± 75 frames at 30 Hz, for 151 frames total).

In Fig. 4*H*, the order of the features is: 1, 2, 17, 18, 3, 4, 5, 19, 6, 20, 7, 21, 8, 22, 9, 23, 10, 24, 11, 12, 13, 25, 14, 15, 16, 26, 27.

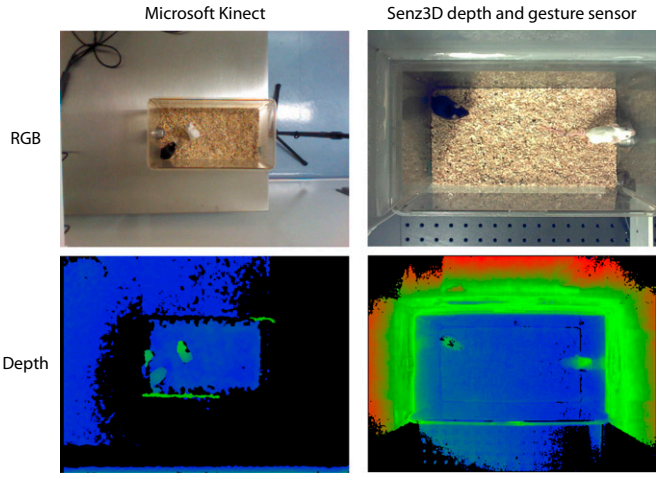


Fig. S1. Comparison between two depth sensors. Representative frames recorded from the Kinect Sensor from Microsoft Corporation (*Left*) and the Senz3D depth and gesture sensor from Creative Technology Ltd. (*Right*).

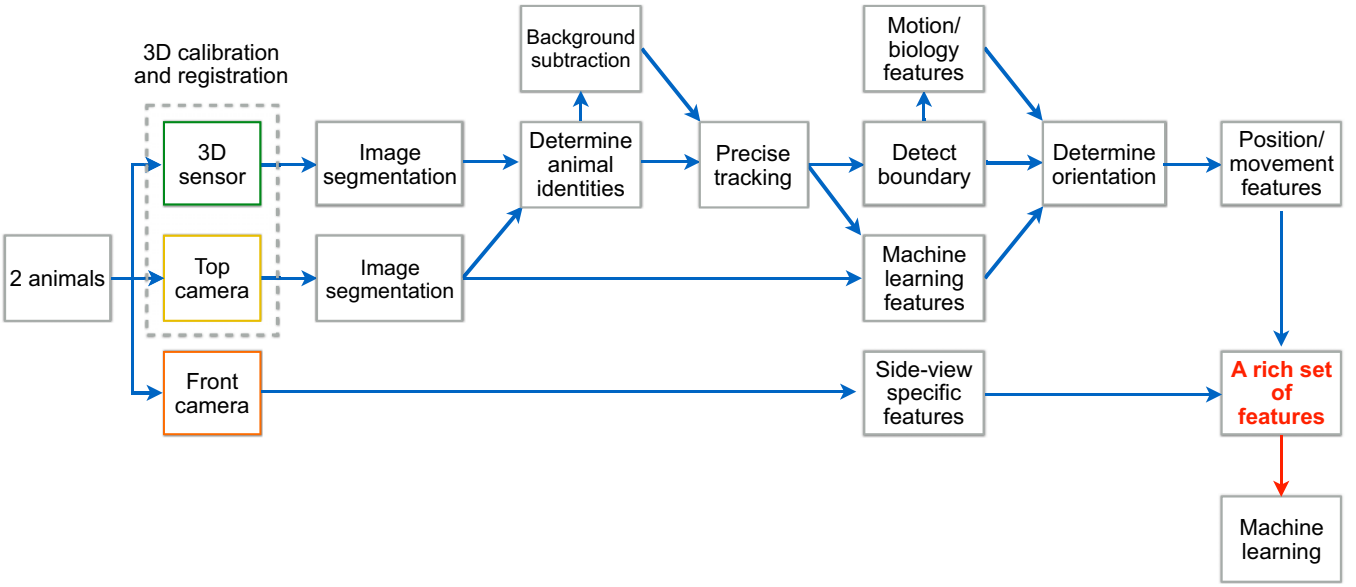


Fig. S2. Detailed workflow. A workflow illustrating the individual steps of the postacquisition image analysis and behavior analysis.

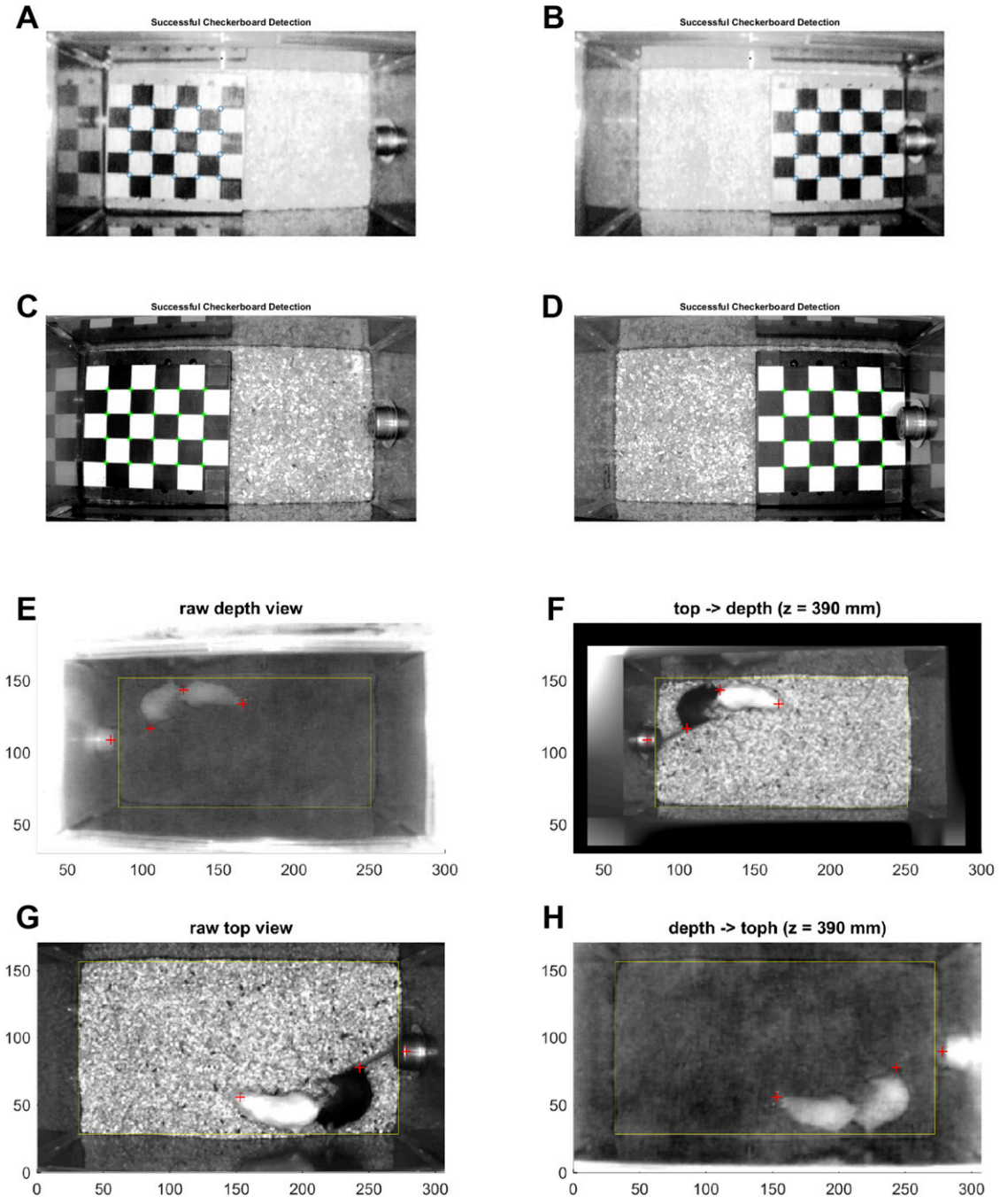


Fig. S3. Registration of depth sensor and top-view camera. (*A–D*) Representative images showing planar checkerboard patterns used to fit parameterized models for each camera. (*A* and *B*) IR images taken by the depth sensor and (*C* and *D*) monochrome images taken by the top-view camera. (*E–H*) Projected video frames in the same coordinated systems. (*E* and *F*) A top-view camera frame (*F*) is projected into the original coordinate of the depth view frame (*E*). (*G* and *H*) A depth-view frame (*H*) is projected into the original coordinate of the top-view camera frame (*G*).

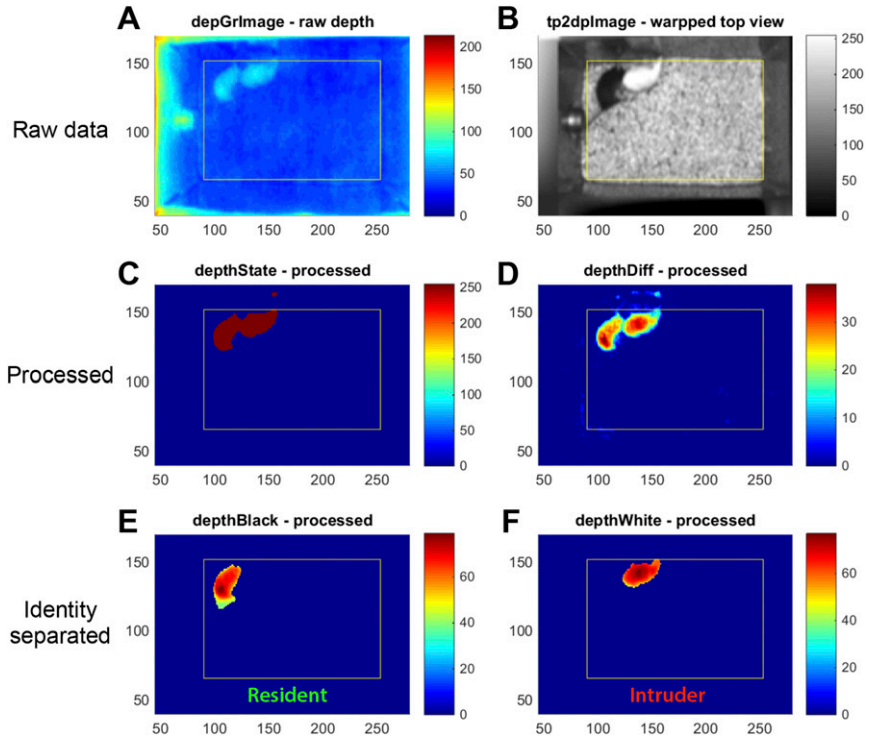


Fig. S4. Imaging processing. (*A*–*F*) Representative images showing raw images taken by the depth sensor (*A*) and the top-view camera (*B*). (*D*) Images with subtracted background. (*C*) Segmented images. (*E* and *F*) Processed images showing the resident (*E*) and the intruder (*F*).

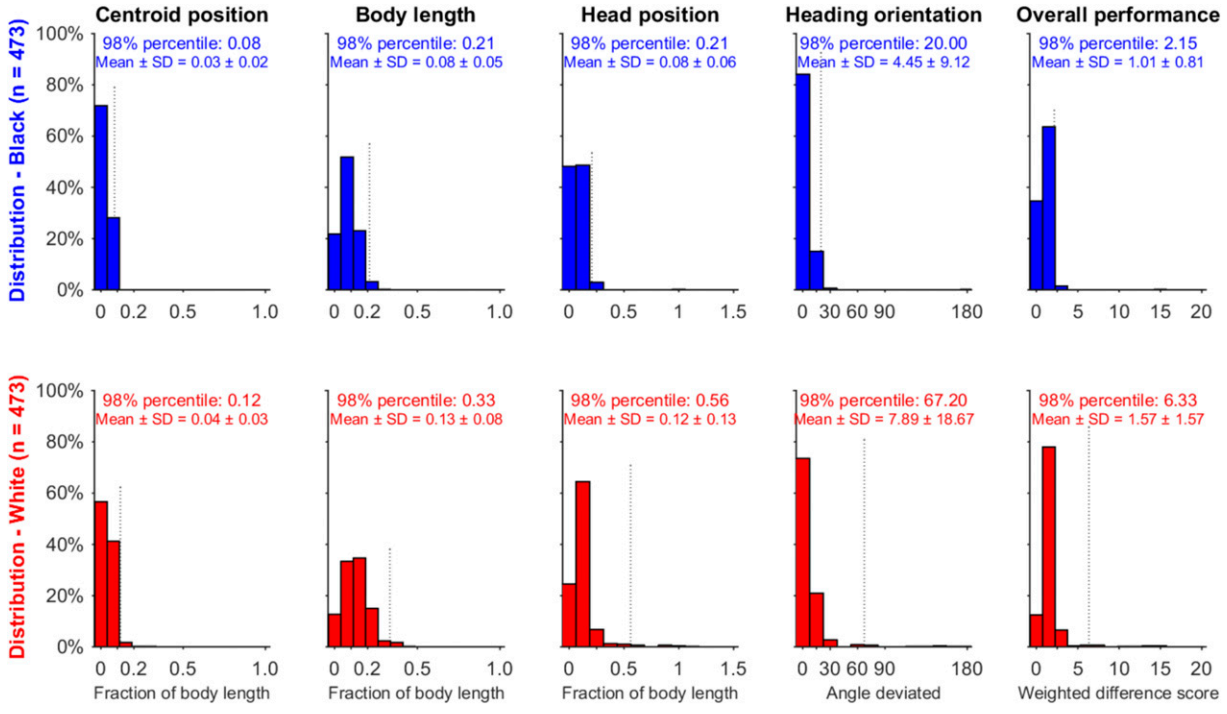


Fig. S5. Comparisons of pose annotations between two independent human observers. Histogram of the difference in each measured parameter among a test set of 400 movie frames annotated independently by two human observers. Dashed line indicates 98th percentiles of the difference for each measurement.

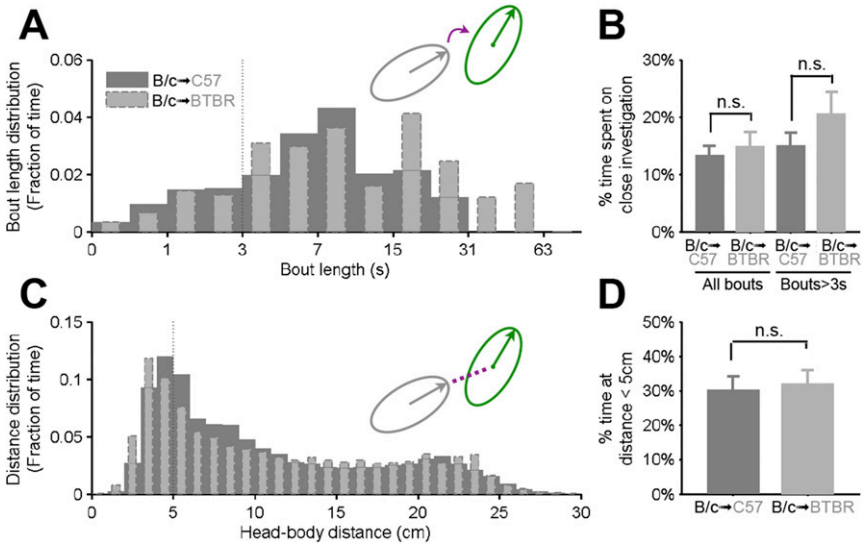


Fig. S6. Social investigation behavior and head–body distance of BALB/c animals toward C57BL/6 or BTBR. (*A*) Histograms of behavioral bout duration (fraction of total time) for social investigation elicited by BALB/c animals toward C57BL/6 or BTBR during the first 5 min of their interactions. (*B*) Percentage of time spent on close investigation during the same behavior sessions. (*C*) Distribution of the distance between the front end of BALB/c animal and the centroid of the subject, BTBR or C57BL/6 (fraction of total time), during the same behavior sessions. (*D*) Percentage of time the front end of the BALB/c animal is within 4 cm from the centroid of the subject (BTBR or C57BL/6) during the same behavior sessions.

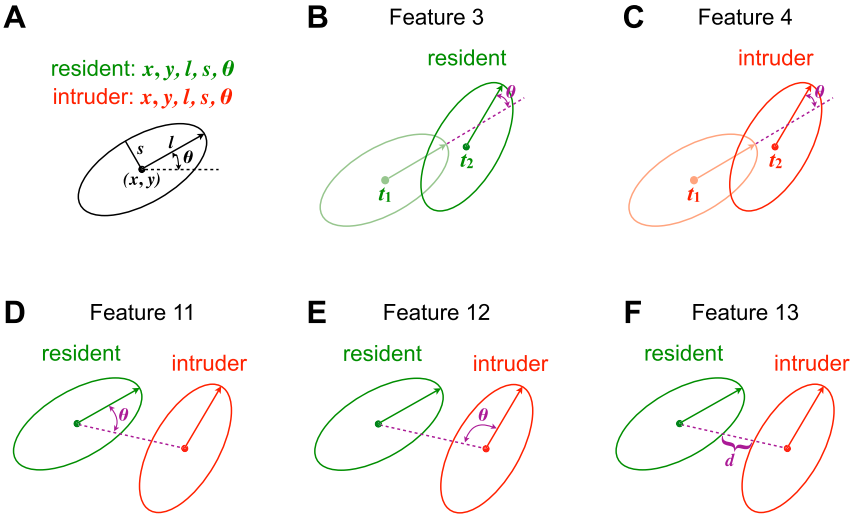
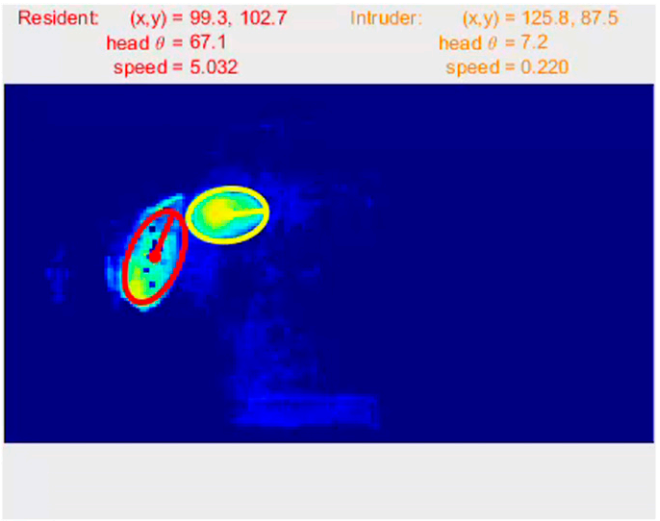


Fig. S7. Schematic illustrating second-order features describing the state of animals in each video frame. (*A*) Ten ellipse parameters (five for each animal). (*B*) Feature 3, describing the change of the body orientation of the resident. (*C*) Feature 4, describing the change of the body orientation of the resident. (*D*) Feature 11, describing the relative angle between body orientation of the resident and the line connecting the centroids of both animals. (*E*) Feature 12, describing the relative angle between body orientation of the resident and the line connecting the centroids of both animals. (*F*) Feature 13, describing the distance between two animals.



Movie S1. A representative movie showing position tracking and postestimation of two animals during social interaction. An ellipse that best fits each animal detected in the segmented depth video frames is used to describe the position, orientation, shape, and scale of the animal.

Movie S1

Movie S2. A representative movie showing the machine learning classification of attack behavior.

Movie S2

Movie S3. A representative movie showing the machine learning classification of close investigation behavior.

Movie S3

Movie S4. A representative movie showing the machine learning classification of mounting behavior.

Movie S4